Article:

Tomova, G.D. orcid.org/0000-0003-1984-8055, Arnold, K.F. orcid.org/0000-0002-0911-5029, Gilthorpe, M.S. orcid.org/0000-0001-8783-7695 et al. (1 more author) (2022) Reply to WC Willett et al. The American Journal of Clinical Nutrition, 116 (2). pp. 609-610. ISSN 0002-9165

https://doi.org/10.1093/ajcn/nqac115

The American Journal of Clinical Nutrition

Reply to DK Tobias et al.
--Manuscript Draft--

Manuscript Number:	AJCN-D-22-00482
Full Title:	Reply to DK Tobias et al.
Article Type:	Invited Reply to Letter to Editor
Corresponding Author:	Georgia D Tomova, BSc MSc University of Leeds Leeds, West Yorkshire UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Leeds
Corresponding Author's Secondary Institution:	
First Author:	Georgia D Tomova, BSc MSc
First Author Secondary Information:	
Order of Authors:	Georgia D Tomova, BSc MSc
	Kellyn F Arnold, BSc MSc PhD
	Mark S Gilthorpe, BSc PhD
	Peter WG Tennant, BSc MSc PhD
Order of Authors Secondary Information:	
Additional Information:	
Question	**Response**
Number of words:	934
Has this manuscript been posted to a preprint server?	No

TITLE PAGE

Reply to DK Tobias et al.

Georgia D Tomova*, Kellyn F Arnold, Mark S Gilthorpe, Peter WG Tennant

Leeds Institute for Data Analytics, University of Leeds, Leeds, LS2 9NL, UK (GDT, KFA, MSG, PWGT)
Faculty of Medicine and Health, University of Leeds, Leeds, LS2 9LU, UK (GDT, MSG, PWGT)
Alan Turing Institute, London, NW1 2DB, UK (GDT, MSG, PWGT)
Faculty of Environment, University of Leeds, Leeds, LS2 9JT, UK (KFA)

*Corresponding Author:
Georgia D Tomova

Address:
Leeds Institute for Data Analytics,
University of Leeds,
Level 11 Worsley Building,
Clarendon Way
Leeds,
LS2 9NL,
UK

Email:
G.D.Tomova@leeds.ac.uk

Phone number:
+447874310249

Word count: 934/1000

Funding: All authors (GDT, KFA, MSG, PWGT) were supported by The Alan Turing Institute (EP/N510129/1) at the time of the original work. The Alan Turing Institute had no influence on this reply.

We are pleased by Tobias et al.'s interest in our recent paper which explores adjustment for energy intake from a causal inference perspective (1). We are aware that some of our conclusions contrast with previous recommendations, which we attribute to the additional insights provided by our use of a contemporary causal inference framework and our novel simulations.

To begin, we strongly agree with the benefits of considering the target trial framework or the equivalent 'control feeding study' when planning and/or evaluating observational studies in nutrition. Unfortunately, while the estimands of interest in an observational study and the corresponding target trial may be the same, the appropriate modelling approach is not always so intuitive. This is shown clearly in our original paper, where we demonstrate that different models can produce different estimates for the same estimand. We respond below to the specific concerns outlined by Tobias et al.

1. **Partition models.** We agree with Tobias et al. that most nutrition research will be interested in isocaloric scenarios. We disagree, however, that an energy partition model that includes all major macronutrients (i.e., the all-components model) is not isocaloric. As we explain in our original paper (1) and elsewhere (2), nutritional data is *compositional*, with the total being mathematically determined by its components. A model containing all components of energy intake therefore fundamentally controls for the total, making it necessarily isocaloric. While the coefficients from such a model do not directly represent relative causal effects, they nevertheless relate to a world where all calories are fixed. The means and SDs for the desired relative causal effects can be easily obtained using, respectively, linear combinations of the model coefficients and the delta method (3); any further distributional properties of interest can be estimated by bootstrapping. Since the standard and all-components models both contain the same exposure, other considerations (such as normality and correlations with confounders) can be examined identically, regardless of the model used.

2. **Standard energy-adjusted models.** Again, we agree with Tobias et al. that the standard model will experience the most problems with a poorly specified estimand. We are clear however that these problems include genuine statistical bias (i.e., a systematic difference between the true and observed coefficients), rather than simply reflecting poor specification. As explained in our paper, this bias occurs for two reasons; 1) fewer components in the model leads to greater residual confounding, and 2) more components summarised by the total energy variable yields greater scope for 'composite variable bias'. In a well-defined substitution, the standard model would contain the exposure, total energy intake, and all remaining components except that being substituted. This is mathematically identical to the all-components model and will perform equally well, with the only difference being that the coefficient obtained describes a relative rather than total causal effect.

3. **Nutrient densities.** We were particularly surprised by Tobias et al.'s defence of the nutrient density model given their past recognition of the problems with this model (4). We agree that it is reasonable to hypothesize dietary composition interventions expressed as percentages; however, we profoundly disagree that using ratio variables in a regression model will accurately estimate these estimands of interest. The problem of ratio variables when used in regression or correlation has been repeatedly described since 1897 by authors including Pearson (5), Fisher (6), and Neyman (7). As best put by Willett and Stampfer themselves in 1986, *"The basic principle involved is that dividing by a variable does not necessarily remove or "control" for the effect of that variable. [...] Since a nutrient density variable contains the inverse of caloric intake as a component, nutrient densities will tend to be associated with disease in the opposite direction to total caloric intake"* (4). Although additional adjustment for total energy in the *multivariable* nutrient density model is advocated to recover the otherwise obscure association, some bias still remains, as demonstrated in our original

paper. This is because adjusting for total energy is not mathematically identical to adjusting for its reciprocal.

4. **Measurement error.** We agree that measurement error is an extremely important concern in nutrition research and should inform the appropriate modelling strategy. However, contrary to Tobias et al., we believe that the presence of measurement error only further strengthens the argument for using the all-components model. Unless the sample size is limited, adjustment for all components should in fact offer the maximum information about the size and nature of any correlated error structure and therefore return the most accurate coefficients for all model variables. Simply adjusting for total energy intake does not provide as much information, and therefore yields more residual measurement error, in addition to the problems of residual confounding and composite variable bias. Dividing by total energy intake (i.e. the nutrient density model) is not advocated under any circumstances, for the reasons outlined above. We will be exploring and explaining the role of measurement error in future work. In the meantime, we have added an updated simulation to the online repository containing the original analytical code that considers various measurement error scenarios. As with the original paper, this shows that the all-components model obtains the most accurate estimates of all originally estimated causal effects.

In summary, we believe that the all-components model offers the most transparent, flexible, and accurate approach to adjusting for energy intake in observational nutrition studies. As demonstrated by our simulations, it is the most effective model for reducing confounding bias, composite variable bias, and measurement error. With increasing awareness of these issues, and increasing adoption of contemporary causal inference methods, we are hopeful for more robust and meaningful causal effect estimates in future nutritional epidemiology.

ACKNOWLEDGMENTS

Authors' responsibilities: GDT and PWGT wrote the reply. All authors contributed to revising the reply and approved the final version before submission.

Conflicts of interest: PWGT and MSG are both directors of Causal Insights Ltd, which sells causal inference research and training; the company and the authors may therefore benefit from any study that demonstrates the value of causal inference methods. GDT and KFA report no conflicts of interest.

REFERENCES

1. Tomova GD, Arnold KF, Gilthorpe MS, Tennant PWG. Adjustment for energy intake in nutritional research: a causal inference perspective. Am J Clin Nutr. Jan 11 2022;115(1):189-198. doi:10.1093/ajcn/nqab266

2. Arnold KF, Berrie L, Tennant PWG, Gilthorpe MS. A causal inference perspective on compositional data. Int J Epidemiol. Aug 1 2020; 49(4):1307-13. doi: 10.1093/ije/dyaa021

3. Oehlert GW. A Note on the Delta Method. Am Stat. 1992;46(1): 27-9, doi: 10.1080/00031305.1992.10475842

4. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am J Epidemiol. Jul 1 1986;124(1):17–27. doi: 10.1093/oxfordjournals.aje.a114366

5. Pearson K. Mathematical contributions to the theory of evolution.—On a form of spurious correlation which may arise when indices are used in the measurement of organs. Proc R Soc Lond. Jan 1 1987;60(359-67):489-98. doi: 10.1098/rspl.1896.0076

6. Fisher RA. The analysis of covariance method for the relation between a part and the whole. Biometrics. Jun 1947;3(2):65-8

7. Neyman J. Lectures and Conferences on Mathematical Statistics and Probability. Second Edi. Nature. Washington: US Department of Agriculture. 1952.